This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.

What [redacted] Should Be [redacted] Classified?

A FRAMEWORK WITH APPLICATION TO THE GLOBAL FORCE MANAGEMENT DATA INITIATIVE

Martin C. Libicki | Brian A. Jackson
David R. Frelinger | Beth E. Lachman
Cesse Ip | Nidhi Kalra

Prepared for the Joint Staff J-8

NATIONAL DEFENSE RESEARCH INSTITUTE

This research was prepared for the Joint Staff Director for Force Structure, Resource, and Assessment (J-8) and conducted within the Acquisition and Technology Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community under Contract W74V8H-06-C-0002.

Library of Congress Control Number: 2010940485

ISBN: 978-0-8330-5001-4

Published 2010 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Summary

DoD frequently needs detailed force structure data to conduct operational planning and budget programming. Each service maintains these data but in different formats, stored on different systems established at different times, and with service-unique labeling that makes aggregating information across the military services difficult, time consuming, and error prone. To address this problem, DoD has initiated GFM DI.[1] Its objective is to standardize formats and protocols so as to ease the process of collecting and aggregating service data for department-level analysis. The great bulk of the data involved are unclassified and therefore accessible by means of DoD's Nonsecure Internet Protocol Router Network (NIPRNET), an unclassified system. Such systems lack security features (e.g., routine content encryption, virtual air-gapping) of the Secret Internet Protocol Router Network (SIPRNET). Because potential adversaries have repeatedly taken data from the NIPRnet, which, broadly speaking, is Internet-accessible, the question has been raised as to whether detailed force-structure data should be stored on SIPRNET exclusively.

The Joint Staff asked NDRI to examine this issue and make recommendations about the need to classify GFM DI information. We addressed this question in two steps. The first was to analyze the reasons for classifying information in general with an eye to distilling some broad criteria that could guide subsequent analysis. The second was to apply the criteria to the data covered by GFM DI.

[1] GFM DI is not a database but a set of standards and connectivity protocols that facilitates the sharing of information stored in various virtual and physical locations.

Why Classify Information?

The most important reason to classify information (and the reason most relevant to GFM DI) is the belief that, if potential adversaries get hold of it, they can use it to undermine U.S. national security. Guidelines exist for determining what national security information should be protected and at what classification level in a tiered system. Assignment to a given level depends on an assessment of how much damage would result if an adversary got the information (e.g., "serious damage" or "exceptionally grave damage").

Some classification decisions—e.g., protecting the identity of a covert intelligence source—are straightforward and prompt little controversy. But other cases are less clear. When does the loss of control over a piece of information lead to "serious damage" or, worse, "exceptionally grave damage"? Without a clear answer, the default decision may be that the information is classified, perhaps overclassified. Classification imposes costs, and these are not just administrative. For one thing, it makes doing business within or among government agencies more difficult. For another, the public has less information about its government.

The "precautionary approach" simply assumes that secrecy confers a benefit and therefore ignores such costs. This avoidance, however, is clearly inadequate for policymaking, which normally requires comparing costs and benefits—understood to include factors that cannot be monetized—to one another.

Thus, to get a handle on the potential damage from adversary gaining specific information, we generated a set of four basic criteria for assessing whether the classification of a particular piece of information has any value:

1. Does classification decrease the amount of information going to potential state and nonstate adversaries?
2. Does the additional information adversaries would have if it is not classified affect what adversaries know (and are such changes meaningful and helpful in the sense that the additional

information moves them closer to, rather than farther from, the truth)?
3. How likely is this change in knowledge to affect possible adversary decisions (and again, does it do so in ways that help the adversary)?
4. Would the decisions the adversary makes based on such knowledge damage U.S. national security?

Only if the failure to classify a piece of information means that an adversary is more likely to get it *and* if having it changes the adversary's estimate of a key piece of knowledge *and* if the change in knowledge alters a decision (or the probability of a decision) *and* if this decision is adverse to the United States would any case exist for classifying it—and then only if the costs of classification, broadly understood, are not greater. If classification yields no measurable benefit, there is no justification for it even if the *costs* of classification are zero, which they never are. In principle, knowledge is power, but not all knowledge is equally powerful.

GFM DI: What It Is, and What Is New About It

GFM DI deals with force-structure information: types of military units and the people who staff them. More specifically, it deals with authorized forces, those detailed in military authorization documents. These forces differ in number from on-hand forces, which are typically those Congress has agreed to pay for, and ready forces, which are those that can actually deploy at any given moment. In the current U.S. wartime posture, both categories are typically smaller than authorized forces.

As mentioned, GFM DI is not a database and does not necessarily require creation of new databases. However, it does mandate that the military services provide a minimum set of essential data about their force elements. The GFM DI data dictionary defines the legitimate values for these required data elements. GFM DI helps integrate service-authorized force management data by allowing users to access what used to be scattered heterogeneous information sources as if they

were one coherent database. However, from the classification standpoint, it is *not* a single database, and dealing with security concerns requires identifying security issues related to either the broader sharing of data or to the aggregation of different data types as a result of the initiative.

What Security Risks Does GFM DI Pose?

The security concerns raised about GFM DI rest on the potential for adversaries—states and nonstate actors alike—to use its data. We framed these concerns in terms of the following three questions:

- Will GFM DI provide adversaries information about the U.S. force structure that they would not otherwise have?
- Will GFM DI make it easier for adversaries to confirm information that they may know already about U.S. forces?
- Will GFM DI's aggregation capabilities create security concerns?

For each question, we examined how the changes that GFM DI would require or possibly induce would affect the access potential adversaries might have to such data. We then examined the data and asked, for each concern, whether the classification or other restrictions on it were supported, based on our four basic criteria.

Having laid out a systematic process for examining those security concerns and determined what the security benefits would be from classification, we analyzed the minimum data set and **found no good reason to classify GFM DI as a whole.** Concerns about the standardization, mandated generation of the minimum data set, and broader utilization of force-structure data appear largely unfounded. In considering how much of an overall picture of U.S. force structure adversaries might gain, we noted that the change was from a status quo of many alternative sources of information to slightly better data, a limited decisionmaking advantage. In the end, the concerns generally failed at least one criterion for considering classification.

Two concerns cannot be dismissed, however. First, within the requirements GFM DI imposes on data providers—that they conform to the GFM DI data dictionary and provide the minimum data set might be revealed about a sensitive unit or platform based on the characteristics of the billets associated with it. Linking GFM DI data to personnel databases (external to GFM DI)—the linkage of *individuals* to billets or units—may also create problems.

Data providers have the flexibility to make data beyond the minimum data set available to GFM DI subscribers. Depending on what data are added to service data sets, adversaries might get access to sensitive data not readily available elsewhere. The main example of this possibility was the use of additional fields to add information about individual force structure components. Although flexibility in building force structure databases and customizing them as needs evolve may benefit DoD, the new data they support—whose extent and nature cannot be predicted now—may create security concerns.

Such concerns are not unique to GFM DI (e.g., sensitive information can be inadvertently released in unclassified email). As a result, no good case could be made that their existence mandates that significant parts of the data covered by GFM DI be classified. The potential security concerns raised can be addressed with better tailored recommendations.

Recommendations

First, caution must be exercised when creating additional data fields or when adding data beyond the minimum data set. Someone (e.g., on or designated by the Joint Staff or the Office of the Secretary of Defense) should periodically scan GFM DI to look for information that should not be there.

Second, the list of displayed fields and the allowable attributes for certain data elements may need to be trimmed. For instance, listing the required security classification of billets may reveal information about individuals that may make them targets for recruitment by adversary intelligence organizations. Similarly, some military occupa-

tional codes are inherently sensitive; a unit with an unexpected number of such billets calls attention to itself as having unexpected missions.

Third, information on units, platforms, or activities that now guard their security through obscurity may have to be classified. Manipulating data is getting easier over time, and people routinely disclose information in a myriad of intentional and unintentional ways, making things simpler for an opponent. The half-life of such tactics is short—with or without GFM DI—and prudent planners should anticipate as much and adjust accordingly.

Fourth, the interaction between GFM DI and personnel databases (external to GFM DI) needs further study. Technically, this issue is outside the purview of GFM DI developers but should, nevertheless, be examined. The ability to link a person to a billet and a billet to a unit may reveal a great deal more about the unit than would GFM DI data alone. The ability to link persons to each other (by linking person to billet to unit to billet to person) allows potential adversaries to conduct a great deal of social network analysis.